

# PDR RID Report

Date Last Modified 7/11/95

Originator Richard Rood

Phone No 301-286-8203

Organization GSFC

E Mail Address rood@dao.gsfc.nasa.gov

Document

RID ID PDR 357

Review SDPS

Originator Ref

Priority 1

Section

Page

Figure Table

Category Name Segment-level

Actionee Project (Kobler) & HAIS

Sub Category Capacity

Subject What keeps these Classic Failures from happening again?

## Description of Problem or Suggestion:

The following are classic problems of NASA data systems:

- 1) Not enough disk space for staging and short-term (less than 30 days) use. Seems especially important where data is sitting and waiting for a healthy network.
- 2) Large and small files: heterogeneous file sizes are the bane of current systems. They introduce unreliability and severely compromise performance by raising "latency."
- 3) Browse and graphics: often of high interest to long-term users and developers, of little interest to near-term users. End up consuming a lot of the data service resources and compromise data delivery, etc. This is often because these were developed first, relying on new machines for their real implementation. Budget and procurement prohibit or delay new machines. Hence, not enough resources.

## Originator's Recommendation

How are these classic failures being avoided? How are the successes of the DAACs being incorporated? What are the paradigm shifts that allows us to avoid them? Are they proven in appropriate benchmarks?

NOTE: The first two items should be answered by HAIS, while the third should be answered by the Project.

GSFC Response by: Ben Kobler

GSFC Response Date 5/5/95

### PROBLEM #3:

We are avoiding classical problems of not having enough resources for browse and graphics by

- 1) working to develop a realistic user model to determine browse and graphics requirements,
- 2) developing an architecture that allows for extensibility, evolvability, and separation of browse/graphics and data delivery,
- 3) planning for the procurement of appropriate hardware to meet browse and graphics requirements.

HAIS Response by: Suhrstedt

HAIS Schedule

HAIS R. E. Richardson

HAIS Response Date 6/30/95

### PROBLEM #1:

We are avoiding classic problems of not having enough staging disk by 1) developing a hardware and software architecture which allows for separation and isolation of differing requirements for working storage, 2) where appropriate, not using classical data movement heuristics for "caching" files to working storage, 3) incorporating results of extensive data modeling, user modeling, and technology analysis to size working storage appropriately.

The data migration and management strategies described in Section 7.4 and 7.7 of DID 305 reflect past and ongoing performance and user modeling efforts and trades and analyses. (Key trades conducted were "Network Attached Storage" White Paper and the "Network Attached Storage Technologies" Study.) These efforts have examined the anticipated data access patterns by ECS users and assessed their impacts upon ECS operations. The capacity of Working Storage for Release-A was determined based on the Processing Subsystem's predicted needs for data pre-staging and production volumes and patterns, staging needs for electronic distribution and the minimum FSMS staging requirements by the Data Server. This capacity sizing reported during PDR, due to a lack of complete data at the time, was not comprehensive. Furthermore, production staging and distribution staging within the DataServer are not separate areas; they are in the same pool of disk storage. Further sizing work planned for CDR will include analysis of Release-A & B Reprocessing and QA loads and use of the system level performance model presently being used to analyze the AHWGP information.

The Network Attached Storage (NAS) Prototype, planned for completion prior to CDR, also supports the sizing of staging disk. Successful implementation of NAS will serve to reduce overall staging disk requirements.

(problems #2 and 3 on next page)

Date of Problem #2/95

Page: 1

Official RID Report

It is true that heterogeneous file sizes pose performance problems for today's data information systems. This is especially true

# PDR RID Report

(problems #2 and 3 on next page)

## PROBLEM #2:

It is true that heterogeneous file sizes pose performance problems for today's data information systems. This is especially true for those systems like ECS which must handle extremely large files as well as files which are also much smaller in size. To avoid this problem ECS has considered the adoption of data storage policies which allow storage and management of files based on their size and anticipated access patterns. For instance, files which are accessed frequently should be stored on devices which allow quick access and large files should be stored on devices that support streaming data at high transfer rates. To support data storage and management design, User Model analyses are being conducted to determine the frequency of demand of various products based on the recent EOSDIS Product Use Survey. Also, the results from the Ad Hoc Working Group on Consumers will provide additional information of IDS teams' needs for various products.

Due to the extremely large volume of data that must be stored by ECS, the choice of data storage devices is limited and will be based upon tape technology which has high density and slow access times. Retrieval of ECS data however, will also involve the use of low density but quick access disk devices. (See section 7.7 of DID 305, the Working Storage HWC)

## PROBLEM #3:

Refer to the GSFC response for an answer to this item.

---

<b>Status</b>	<b>Closed</b>	<b>Date Closed</b>	<b>7/11/95</b>	<b>Sponsor</b>	<b>Szczur</b>
---------------	---------------	--------------------	----------------	----------------	---------------

\*\*\*\*\* Attachment if any \*\*\*\*\*

---